

Managing the Evolution of SAS® Programming

Carey Smoak, Portola Pharmaceuticals, South San Francisco, CA

ABSTRACT

Over the past 30+ years, I have seen many innovations in SAS as a software tool. While SAS software has become more and more innovative, the SAS programming profession in the pharmaceutical / biotechnology industry has both evolved and de-evolved in some important ways. The evolution is mainly due to advances in SAS as a software tool. Thus the SAS programmer is challenged to keep up with new innovations in the SAS software. At the same time, the daily work of a SAS programmer may be de-evolving due to working in a specification driven environment and standardization (CDISC and macros). Standardization and specifications can be beneficial, but SAS programmers may become bored in environments that are highly standardized. The recent growth in the data scientist profession may offer a way for SAS programmers in the pharmaceutical / biotechnology industry to grow their careers and take on more daily tasks. This paper will present some ideas for growing the SAS programming profession.

INTRODUCTION

The first version of SAS that I used was SAS79 (prior to version 5, SAS released versions yearly, so SAS79 was released in 1979). There have been many innovations in SAS since I first began using SAS. SAS has gone from being primarily statistical software to an array of software products and solutions. Just one example is the ability to integrate SAS with Microsoft Office products to generate real-time monitoring reports (Truong, Smoak 2014; Smoak 2013; Smoak 2014). While SAS software has innovated greatly in the 30+ years that I have been using SAS, the actual industries which use SAS has also been innovating. Recently “big data” and “data science” have become big buzzwords in the statistics profession (Kincaid 2013).

While SAS has greatly innovated over the years, it is my opinion that the SAS programming profession has both evolved and de-evolved over the past 10 years or so. From a managerial perspective, the purpose of a manager of SAS programming is to assist their staff in being successful. A SAS programmer will be highly successful when their skills are growing and they are properly motivated to perform to the best of their abilities (Smoak 2009). Thus a manager of SAS programming must encourage their staff to keep up with innovations SAS programming to keep growing as a SAS programmer. For example, the recent innovations in SGPLOT make the production of graphs much easier than the tradition GPLOT procedure.

On the other hand standardization of processes (e.g., CDISC and macros) in the pharmaceutical industry may hinder innovation for certain SAS programmers. For example, work in a highly standardized environment may mean that the SAS programmer mainly passes parameters into macros and does not really know what the macro is doing. For several years, I have been doing telephone screening interviews of candidates for SAS programming positions. In general, I would estimate that about 50% of the SAS programmers that I have screened cannot completely answer my data step screening questions. Usually, I ask just three data step questions in a screening interview. This may be a generalization, but a trend I have observed is that SAS programmers who work in a highly macroitized environment tend to be the ones who cannot fully answer my data step screening questions. Therefore, my question for this paper is whether or not standardization can stymie innovation for a SAS programmer.

Therefore, the dilemma is how does a SAS programmer who is in a highly standardized environment keep motivated to grow in their skills and stay motivated? The answer may partially lie with redefining the role of a SAS programmer from a spec-driven programmer to more of an analyst.

SAS AND THE FDA

First, while the FDA does currently require that version 5 SAS transport files be submitted (FDA 2014), there is no general requirement that SAS be used to generate summary tables, listings and figures for an FDA submission. So, why has the pharmaceutical industry widely adopted SAS as the standard for programming of analysis datasets, summary tables, listings and figures? One important reason is that SAS can be validated and validation is important to the FDA (Benze 2005). While validation is very important, does that necessarily mean that software products like R cannot be used in an FDA submission? In fact, there is evidence that the FDA has accepted with the use of R in FDA submissions (Smith 2012).

DATA SCIENCE

Data scientist is a job title that has grown in popularity in recent years. But, what is a data scientist? Perhaps one way to describe a data scientist is a person who has good programming skills, a good knowledge of math / statistics and substantial expertise in a particular industry (Kincaid 2013). While this general description may sound very much like a SAS programmer in the pharmaceutical / biotechnology industry, there are substantial differences between a data scientist and a SAS programmer in the pharmaceutical / biotechnology industry.

According to Kincaid (2013), there are 5 areas which distinguish a data scientist:

- Scripting languages: Perl, JavaScript, Python, Ruby
- Programming languages: C#, C/C++, Java
- Database and Big Data languages: HiveQL, MySQL, NoSQL, SQL
- Statistical Languages: R, SAS, Matlab, Weka, Tableau

Not all of these skills may be required of a data scientist for a particular job. Therefore, this list is intended to show the range of skills of the data science profession.

What further delineates a data scientist from a SAS programmer in the pharmaceutical / biotech industry? The data scientist is responsible for producing solutions to business problems (Kincaid 2013). In contrast, SAS programmers in the pharmaceutical / biotechnology industries are primarily specification-driven to produce analysis datasets and TLFs (Table, Listings and Figures). Thus the question must be asked as to whether or not SAS programmers can take on more responsibility by adding data scientist skills to become a partner in solving problems in the pharmaceutical / biotechnology industry. This proposes the question as to whether or not this increased role for SAS programmers would impinge on the role of biostatisticians or be of assistance to them? My personal observation is that there seems to be a lack of biostatisticians in the pharmaceutical / biotechnology industry. Thus the increased role of SAS programming may actually benefit this industry.

PUTTING IT TOGETHER: THE FUTURE OF SAS PROGRAMMING IN THE PHARMACEUTICAL INDUSTRY

RESPONSIBILITIES OF A SAS PROGRAMMER

The typical job description of a SAS programmer in the pharmaceutical / biotechnology industry may include:

- Good knowledge of SAS (including macros) and CDISC
- Good communication skills
- Write and execute SAS programs for creation of SAS data sets, tables, listings and figures
- Maintain SAS macro library
- Ability to oversee other programmers assigned to a project (for senior/lead SAS programmers)

But what does this translate to on a daily basis for a SAS programmer? Typically, SAS programmers spend their time writing SAS code based upon specifications (which usually come from either a study statistician and/or a senior/lead SAS programmer). Usually, SAS programmers have to work quickly and efficiently to meet deadlines. Clarifying specification may take a lot of back-and-forth communication until the programmer knows exactly what needs to be produced by the SAS program that they are engaged in.

How could the responsibilities of a SAS programmer be different by incorporating data scientist skills? Would the inclusion of these new skills conflict with the work of statisticians?

In my years as both a SAS programmer and a manager of SAS programmers, the issue of specifications for SAS programmers has been a constant issue. Statisticians have a lack of time to produce the level of specifications needed by SAS programmers. Thus in some companies SAS programmers have taken on creating specifications such as SDTM. However, the real issue is that many SAS programmers do not have the mindset that they can produce specifications. What if expectations for SAS programmers changed so that they were expected to be participants in the specification process? From my experience, statisticians would welcome the partnership of SAS programmers being co-owners of the specifications.

Thus the level of back-and-forth communication between statisticians and SAS programmers could be elevated by expecting more input from the SAS programmer. Thus the SAS programmer role would change from clarifying specifications to being a partner in the specification process. This elevation of expectations for SAS programmers will be illustrated in the following sections.

SAMPLE STUDY

The typical work of a SAS programmer is to produce safety and efficacy data sets which will be used to generate TLFs to support the study claims. With the advent of CDISC standards such as CDASH, SDTM and ADaM, the expectation is that SAS programming will become more automated. Certainly, standards can produce efficiencies. However, not all studies are exactly alike. The study statistician has to be involved in the proper design of a study. Therefore, analysis of individual studies will vary.

As an example, let's take a pivotal phase III study. The primary objective may be "is Drug A better than Placebo?" The study statistician will be involved in the proper design of this study. Typically, the CRFs will be designed to collect the correct data to support the claim(s) of the study. Following CDASH standards for data collection will help with the conversion of the data from the database to SDTM. The objectives of the study will guide the ADaM data sets to be created and the TLFs to be produced.

Typically, in this process, the statisticians drive the specifications for the SAS programmer to produce the ADaM data sets and the TLFs. How would this change if the expectations of the SAS programmer were such that they were seen as partners in the specification process? Generally, SAS programmers are good at knowing how to produce output which looks good. In other words, many statistical programmers enjoy spending time to make output look pretty. What if the generation of the TLF shells was a partnership and the SAS programmer participated from the beginning in designing output which looked good? This could cut down on the amount of time in clarifying specifications by seeing the initial specifications as a partnership between the study statistician and the SAS programmer.

THE "REAL LIFE" OF A SAS PROGRAMMER

SAS programmers are typically busy writing code to meet deadlines. So, why would a SAS programmer want to take on data science skills to add to their repertoire? There could be a variety of motivations: learning new skills, career advancement, process improvement and so on.

What challenges would a SAS programmer face by incorporating data science skills? One challenge could be balancing the passion and curiosity of a scientist who explores the data in new ways and even adds new data to the analysis, with the need of the company to capitalize on the standardization so that the time to market is reduced. Thus the SAS programmer would truly be a partner in the analysis of clinical trial data. Thus the SAS programmer would be more like an analyst/data scientist. In the end, this would be a paradigm shift for the industry and may take time for acceptance of this new role.

The goal of this paper has been to present the idea that expectations of the SAS programmer may shift in the future. Is this a radical idea? Not really, some companies are already changing the role of the SAS programmer to be a partner with statisticians. Thus this is an ideal time to open and expand the discussion on the role of the SAS programmer. The role of a data scientist may be appropriate to add to this discussion. In other words, by examining the characteristics of a data scientist we may discover ways to help the role of the SAS programmer to evolve. Thus the "real life" of a SAS programmer could change from being specification driven to being more of a partner in the analysis of the clinical trial data.

WRAPPING IT UP

As indicated earlier in this paper, the question of the evolving and de-evolving nature of SAS programming as a profession in the pharmaceutical industry was raised. It is not an uncommon experience for a SAS programmer to be excited about a new procedure in SAS only to be told that it would mean re-validating already validated macros. Thus using new SGPLOT procedures instead of old GPLOT procedures may be inhibited. Also the use of software other than SAS (e.g., R) may be inhibited because of the perception that only SAS can be used in FDA submissions.

Thus a SAS programmer in the pharmaceutical industry may feel stymied in their career. While this paper does not claim any definitive solutions to the problem, it does raise questions about the future of SAS programming in the pharmaceutical industry. Is it possible to transform the SAS programmer into more of an analyst that partners with statisticians? Can the SAS programmer learn skills from the new and exciting data science profession? Can the SAS programmer add other languages like R to their repertoire?

In other words, can a SAS programmer in the pharmaceutical / biotechnology industry benefit from learning more data science skills? My personal opinion is yes. As SAS programming becomes more standardized in the pharmaceutical / biotechnology industry, there is definitely room for growth of SAS programmers to assume more responsibility and not to merely be specification-driven. Much of the specification driven work may be automated and, thus, give the SAS programmer more time to do more challenging tasks. I call this idea "making easy things easy," so that one can spend more time on more challenging tasks.

CONCLUSION

While SAS has evolved over recent decades as a software tool, the profession of SAS programming in the pharmaceutical / biotechnology industry is both evolving and de-evolving. The evolution is mainly due to advances in SAS as a software tool and the de-evolving is due to the specification driven programming environment and

standardization (CDISC and macros). Data science is a rapidly growing profession and may offer a solution to grow the careers of SAS programmers in the pharmaceutical / biotechnology industry. Incorporating skills from data science may allow the SAS programmer to do more than is currently being done in the industry. Thus discussions on ways to incorporate data science skills into the skill set of SAS programmers in the pharmaceutical / biotechnology industry is a worthwhile discussion that needs to happen.

ACKNOWLEDGEMENT

I would like to thank Chuck Kincaid (Expersis Business Intelligence and Analytics) for his review of this paper.

REFERENCES

- Benze, Keith. 2005. "Risk-Based Approach to SAS® Program Validation." *Proceedings of the Pharmaceutical SAS User Group 2005 Conference*. Phoenix, AZ. Available at: <http://www.pharmasug.org/2005/FC04.pdf>
- FDA. 2014. "Electronic Regulatory Submissions and Review Helpful Links." Available at: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM085361>
- Kincaid, Chuck. 2013. "How to be a Data Scientist Using SAS®." *Proceedings of the Western Users of SAS Software 2013 Conference*. Las Vegas, NV. Available at: http://www.lexjansen.com/wuss/2013/70_Paper.pdf
- Smith, David. 2012. "FDA: R OK for drug trials." Available at: <http://blog.revolutionanalytics.com/2012/06/fda-r-ok.html>
- Smoak, Carey. 2009. "Empowering SAS® Programmers: The Role of the Manager." *Proceedings of the Pharmaceutical SAS User Group 2014 Conference*. Portland, OR. Available at: <http://www.lexjansen.com/pharmasug/2009/ma/MA03.pdf>
- Smoak, Carey. 2013. "Challenges in Moving to a Multi-tier Server Platform." *Proceedings of the Western Users of SAS Software 2013 Conference*. Las Vegas, NV. Available at: http://www.lexjansen.com/wuss/2013/114_Paper.pdf
- Smoak, Carey. 2014. "Challenges in Moving to a Multi-tier Server Platform." *Proceedings of the Pharmaceutical SAS User Group 2014 Conference*. San Diego, CA. Available at: <http://www.lexjansen.com/pharmasug/2014/AD/PharmaSUG-2014-AD08.pdf>
- Truong, Sy; Smoak, Carey. 2014. "Delivering Dynamic, Real-Time Reports to MS Office Users as Stored Processes." *Proceedings of the Western Users of SAS Software 2014 Conference*. San Jose, CA. Available at: http://www.lexjansen.com/wuss/2014/110_Final_Paper_PDF.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Carey Smoak
Director, Statistical Programming
Portola Pharmaceuticals
270 E. Grand Ave
South San Francisco, CA 94080
csmoak@iportola.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.